
COMMENTARY

Five retracted structure reports: Inverted or incorrect?

BRIAN W. MATTHEWS

Institute of Molecular Biology, Howard Hughes Medical Institute, and Department of Physics, University of Oregon, Eugene, Oregon 97403, USA

Recently, Chang et al. (2006) withdrew five articles that included reports of the structure of the ABC transporter MsbA (Chang and Roth 2001) and the EmrE multidrug transporter in complex with a substrate (Pornillos et al. 2005). The stated reason for the retractions was as follows: "An in-house data reduction program, which was not part of a conventional data processing package, converted the anomalous pairs (I_+ and I_-) to (F_- and F_+), thereby introducing a sign change. As the diffraction data collected for each set of MsbA crystals and for the EmrE crystals were processed with the same program the structures reported... had the wrong hand."

The purpose of the present commentary is to point out that the interconversion of F_+ and F_- does not, in general, lead to an inverted structure. Rather, it leads to a "non-sense" electron density map that has no relation to the true structure or to its mirror image. There are special situations where an inverted electron density map can be obtained, but if such a map is observed, it shows that an error has occurred. The following is an attempt to provide additional background, particularly for the nonexpert.

The increased brilliance of synchrotron X-ray sources, coupled with ever-increasing computing power, has made it possible in favorable cases to collect data and solve structures within hours, if not minutes. At the same time, this has led to the ever-increasing use of "black box" procedures for structure determination. Such procedures are especially powerful when moderate-to-high resolution data are available (better than ~ 2.5 Å), but need to be used with discretion where only poorer quality X-ray data can be measured. As the resolution becomes poorer, electron density maps become harder to interpret, leading to less reliable starting models. Model refinement is also

increasingly difficult. Procedures in which multiple copies of the protein are refined (Chang and Roth 2001) can improve refinement statistics but are of questionable validity with low-resolution data in that they increase what is already a very unfavorable ratio of model parameters to X-ray observations.

The atoms in a crystal can be considered as lying on sets of parallel planes, each set of planes being identified by three integers, h , k , and l . $F(h,k,l)$ gives the amplitude of X-rays scattered from the "front-side" of the (h,k,l) planes and $F(-h,-k,-l)$ the amplitude of scattering from the "back-side" of the same planes. $F(h,k,l)$ and $F(-h,-k,-l)$ are often abbreviated as F_+ and F_- . If a protein crystal contains only light atoms such as hydrogen, oxygen, etc., the amplitudes F_+ and F_- will be identical. If, on the other hand, the protein crystal includes heavy atoms such as mercury or selenium, then F_+ and F_- will differ slightly. This small difference in amplitude occurs because the heavy atoms scatter X-rays "anomalously," i.e., slightly out of phase relative to light atoms. Because it is anomalous scattering that causes F_+ and F_- to differ, they are often referred to as "anomalous pairs."

The traditional method of determining crystal structures of proteins is by the use of a series of isomorphous heavy-atom derivatives. Such derivatives are typically obtained by soaking heavy atoms into pregrown crystals of the native protein or by growing crystals in the presence of heavy-atom-containing compounds. Today, structure determination still requires a form of the crystal that contains one or more heavy atoms. Replacement of methionine with selenomethionine is frequently used to directly incorporate the heavy atom selenium into the protein of interest (Hendrickson et al. 1990; Hendrickson 1991).

Except where exceptionally high-resolution data are available, and "direct methods" can be used, the first step in solving the structure of a protein is to determine the position of the heavy atoms. This is done with the so-called Patterson function, but the solution is always ambiguous. One solution will correspond to the true

Reprint requests to: Brian W. Matthews, Institute of Molecular Biology, Howard Hughes Medical Institute, and Dept. of Physics, 1229 University of Oregon, Eugene, OR 97403-1229, USA; e-mail: brian@uoregon.edu; fax: (541) 346-5870.

Article published online ahead of print. Article and publication date are at <http://www.proteinscience.org/cgi/doi/10.1110/ps.072888607>.

coordinates of the heavy atoms. The second solution will be inverted; i.e., all of the heavy-atom coordinates, (x, y, z) , will be replaced by $(-x, -y, -z)$. Without additional information there is no way to determine which arrangement of the heavy atoms is correct.

Historically, one might go ahead and determine the structure of the protein using just isomorphous replacement measurements (i.e., no anomalous scattering information is included). This was the situation when Kendrew's group determined the structure of myoglobin (Kendrew et al. 1960). Suppose, by chance, you choose the correct set of heavy-atom coordinates and calculate an electron density map for the protein. This map will show the true structure. Suppose, however, you choose the inverted heavy-atom positions and calculate an alternative electron density map. This map will be inverted relative to the correct one and will give a mirror-image protein structure (e.g., all of the α -helices will be left-handed and all of the amino acids will be D rather than L). The investigator compares the two alternative maps, keeps the one with the right-handed

helices, and ignores the other. The key point in this situation is that no anomalous scattering information has been included. Because Chang and Roth (2001) were using anomalous scattering data, a different result is expected (see below).

In the case where anomalous scattering data are included, the Patterson function is still used to locate the heavy atoms. Also, as above, two alternative (correct and inverted) sets of coordinates are obtained for the heavy atoms. The respective sets of heavy-atom coordinates are used to calculate two alternative electron density maps. The difference is that the anomalous scattering data are now included in the determination of the phase angles. In the case where the correct heavy-atom arrangement has been chosen, you obtain a "true" electron density map. In the case where the wrong (inverted) set of heavy-atom coordinates is used, you do not obtain an inverted map. Rather, you obtain a map that consists essentially of noise.

The above "right" and "wrong" calculations using anomalous data are illustrated in Figure 1, A and B, taken

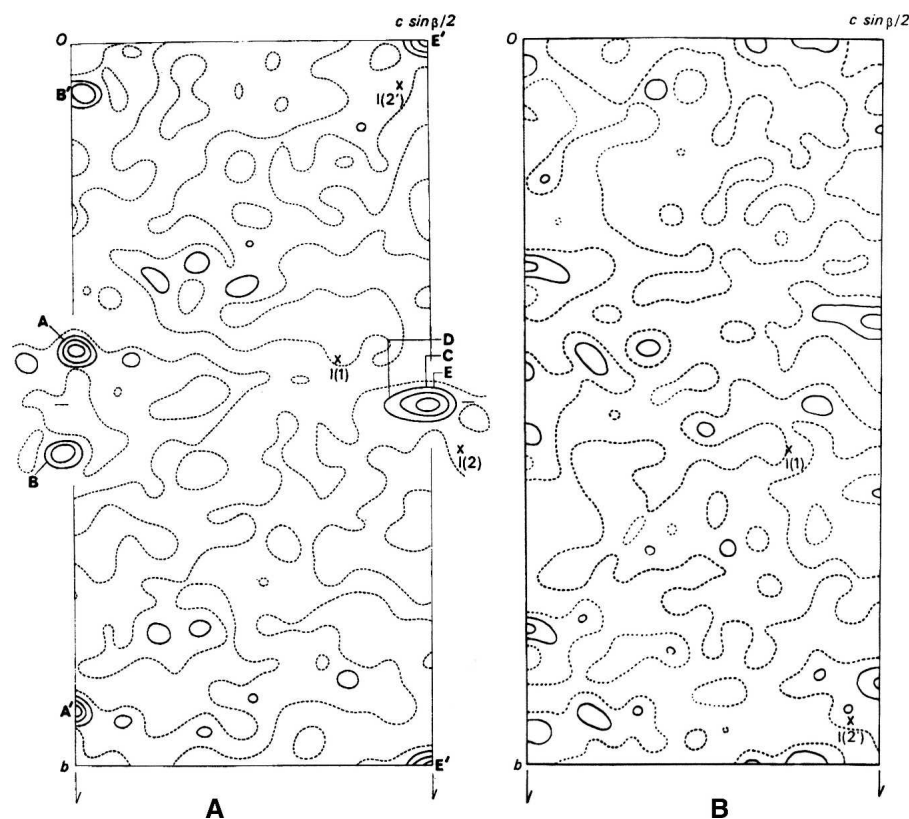


Figure 1. Consequences of the use of "inverted" versus "correct" heavy-atom coordinates on protein phases calculated using isomorphous replacement and anomalous scattering data. As discussed in the text, iodine atoms I(1) and I(2) in one heavy-atom derivative are used to calculate protein phases that in turn are used to locate the heavy atoms A–E in a second heavy-atom derivative. (A) The correct choice for the coordinates of I(1) and I(2) leads to correct phases, which are confirmed by the peaks at the heavy atom sites A–E. Because the electron density is projected on to a plane, atoms C, D, and E overlap. (B) The coordinates of the iodine positions are inverted. The same isomorphous replacement plus anomalous scattering data are used in the alternative phase calculations. The use of the inverted coordinates for I(1) and I(2) leads to "nonsense" phases, and the resultant electron density map consists of noise peaks (from Matthews 1966).

from an early study of α -chymotrypsin (Matthews 1966). In this case, a heavy-atom derivative containing two iodine atoms I(1) and I(2) was used to calculate (approximate) phase angles. These phases were not used to calculate a map for the protein itself, but to calculate a “difference map” that would (hopefully) locate the positions of the different heavy atoms A, B, C, D, and E in a second (PtCl₄) derivative. In Figure 1A, the iodine atoms I(1) and I(2) happen to be the correct choice, and the resultant phases allow one to see clear peaks at the expected positions for PtCl₄ sites A–E in the other derivative. In the calculation shown in Figure 1B, the coordinates of heavy atoms I(1) and I(2) have been inverted. Now the phases are incorrect and no peaks appear for the atoms A–E; i.e., Figure 1B is not the inversion (or mirror image) of Figure 1A. Carrying out two parallel calculations as in Figure 1, A and B, is now sometimes used as a way to determine which of the two heavy-atom arrangements is correct. This, in turn, allows the investigator to obtain a “true” electron density map showing the protein structure with the correct hand.

All of the above assumes that the anomalous pairs have been indexed correctly. If an error is made and they are interchanged, this error has essentially the same consequences as choosing the wrong handedness for the heavy-atom arrangement; i.e., if F₊ and F_− are inadvertently interchanged, the electron density that will be obtained will not be the mirror image of the true electron density, it will be essentially noise. Suppose, however, that the wrong (inverted) coordinates had been chosen for the heavy atoms and F₊ and F_− were interchanged by error. In this situation the resultant electron density map would be inverted relative to the correct map (Table 1). As noted above, in such a map the α -helices would be left-handed and the

chirality of the amino acids inverted. Because such a map had been obtained, it would signal that some sort of error must have occurred (e.g., the interchange of F₊ and F_−).

It might be noted that it should not be difficult to distinguish between a correct and an inverted map. An α -helix is right-handed, whether viewed from the N to the C terminus or vice-versa. Therefore, it is not necessary to know the direction of the polypeptide chain. The differentiation between electron density corresponding to a right-handed and left-handed helix should be apparent even at relatively low resolution. Also, if a protein includes a region of β -sheet, this will almost always have a distinct clockwise twist (Chothia 1973) and would be anticlockwise in an inverted model.

Notwithstanding all of the above, there is still a special situation that needs to be considered and may be relevant to the MsbA structure determination of Chang and Roth (2001). What happens if the arrangement of the heavy atoms has a center of symmetry? In this case the use of the Patterson function to determine the coordinates of the heavy atoms gives an unambiguous solution. It shows the heavy atoms in their correct centrosymmetric arrangement. It is no longer required to differentiate between one heavy atom arrangement that is correct and another that is not. If anomalous scattering data are available, one can calculate phase angles for the protein knowing that the phases should be correct and reveal the true structure. If one had made a mistake and interchanged F₊ and F_−, the resultant electron density would be inverted relative to the true map. As before, if such a map were obtained, it would be a clear indication that some sort of error must have occurred.

As summarized in Table 1, it turns out that the overall conclusion is the same, whether the heavy-atom coordinates

Table 1. How the interchange of the anomalous pairs F₊ and F_− influences protein electron density maps

Method of phase determination	Heavy-atom coordinates	F ₊ and F _− identified correctly?	Resultant electron density map
Isomorphous replacement without anomalous scattering	Correct hand	N/A	Correct
	Inverted hand	N/A	Inverted
	Centrosymmetric	N/A	Correct plus inverted
Isomorphous replacement with anomalous scattering, or multiple wavelength anomalous dispersion (MAD)	Correct hand	Yes	Correct
	Correct hand	No	Noise
	Inverted hand	Yes	Noise
	Inverted hand	No	Inverted
	Centrosymmetric	Yes	Correct
	Centrosymmetric	No	Inverted

The heavy-atom coordinates are determined from the Patterson or related functions. There is an ambiguity between the correct set of heavy-atom sites and another set of sites related by a center of symmetry (i.e., by inversion). Both alternatives need to be considered during the structure determination. If the X-ray data are collected and processed correctly, the anomalous pairs F₊ and F_− will be identified uniquely and there is no requirement to choose between two alternatives. The last column describes the type of electron density map expected depending on whether the coordinates chosen for the heavy atoms have the correct hand, and whether a mistake occurred in the identification of F₊ and F_−. The table also describes the electron density maps that will be obtained if the arrangement of heavy atoms happens to be centrosymmetric. If multiple heavy-atom derivatives are being used, the “centrosymmetric” case applies only if all heavy atoms in all derivatives have a common center of symmetry.

are centrosymmetric or not. If anomalous scattering data are included in the phase determination, and the F_+ , F_- pairs are identified properly, there is no situation in which an inverted electron density map can be obtained. If F_+ and F_- are interchanged, there are situations in which an inverted electron density map could result, but obtaining such a map should immediately signal that an error had occurred.

In the case of the MsbA structure of Chang and Roth (2001), the space group is $P1$, and it appears from Figure 2 of their manuscript that there are two heavy atoms (OsCl_3) in the unit cell. (No osmium sites are included in the three sets of coordinates for MsbA in the Protein Data Bank [1JSQ, 1PF4, and 1Z2R], so the question of centrosymmetry remains open.) If there are just two heavy atoms, and they have equal occupancy, this would be a centrosymmetric arrangement (the center of symmetry being midway between the two OsCl_3 groups). A centrosymmetric arrangement of the heavy atoms, coupled with a mistaken identification of F_+ and F_- , could give rise to an inverted map. If this was not identified as such, and an attempt made to model the density with conventional right-handed helices and L-amino acids, the model would not refine in a satisfactory fashion. Such an "illegitimate" model, as opposed to a truly inverted model, would give rise to abnormally high values of the crystallographic residuals R and R_{free} . Extraordinarily high values of R and R_{free} were obtained for the single model refinement of MsbA but were discounted because of presumed disorder (Chang and Roth 2001). Chang and Roth (2001, Reference 29) also state that "The correct hand of the structure was established by observing the hand of the α -helices..." Where anomalous scattering data are included, as for the MsbA structure determination, the situation should never arise that one would need to choose between the correct map and its inverted image.

In conclusion, we return to the illustration shown in Figure 1. Calculations of this sort, although "old fashioned," remain as one of the most powerful methods to check the reliability of a crystallographic structural analysis, especially when the resolution of the data is limited. As noted above, the electron density maps in Figure 1 are intended to show the locations of a set of PtCl_4 heavy-atom sites. Rather than having to evaluate electron density maps of the whole protein, the investigator has a much simpler task, namely to determine whether the maps reveal the locations of a small number of isolated heavy atoms. Such heavy-atom sites can be easily recognized in maps that are of far lower resolution than those required to interpret the three-dimensional structure of the protein.

In Figure 1, the phase angles for the calculation were obtained from a putative set of iodine heavy-atom sites. In other contexts, the phase angles may be derived from

a presumptive model for the protein structure. If the calculation is "correct," the desired heavy-atom sites will be revealed by distinct positive peaks many standard deviations above background (in a contemporary calculation, three-dimensional data would be employed, giving a much greater signal-to-noise ratio than in Fig. 1A). The absence of distinct positive peaks (as in Fig. 1B) indicates that there is a deficiency either in the calculation or in the data on which the calculation is based.

It appears that Pornillos et al. (2005) used calculations of this type in their analysis of the structure of EmrE. EmrE is a different type of transporter protein, not directly related to MsbA. The structure analysis was based on an arsenic-containing form of the protein in one space group and a selenomethionine-containing form in a different space group. Pornillos et al. (2005, Supplemental material) state that "the close similarity between the two crystal forms allowed us to directly map the SeMet positions to the native structure and also confirm later by anomalous Fourier." Figure 1A of Pornillos et al. (2005) shows "anomalous difference density... (for) As, contoured at 1σ ." Additional peaks of height 4σ are attributed to selenium atoms. Details of the calculation(s) are not given, but at face value the low significance of the arsenic peak, in particular, and also the modest significance of the selenium peaks strongly suggest a deficiency in the overall crystallographic analysis.

Acknowledgments

Helpful comments from Drs. Walt Baase, S. James Remington, Dale Tronrud, Alex Wlodawer, and Zac Wood are greatly appreciated.

References

- Chang, G. and Roth, C.B. 2001. Structure of MsbA from *E. coli*: A homolog of the multidrug resistance ATP binding cassette (ABC) transporters. *Science* **293**: 1793–1800.
- Chang, G., Roth, C.B., Reyes, C.L., Pornillos, O., Chen, Y.-J., and Chen, A. 2006. Retraction: Structure of MsbA from *E. coli*: A homolog of the multidrug resistance ATP binding cassette (ABC) transporters. *Science* **314**: 1875.
- Chothia, C. 1973. Conformation of twisted β -pleated sheets in proteins. *J. Mol. Biol.* **75**: 295–302.
- Hendrickson, W.A. 1991. Determination of macromolecular structures from anomalous diffraction of synchrotron radiation. *Science* **254**: 51–58.
- Hendrickson, W.A., Horton, J.R., and LeMaster, D.M. 1990. Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): A vehicle for direct determination of three-dimensional structure. *EMBO J.* **9**: 1665–1672.
- Kendrew, J.C., Dickerson, R.E., Strandberg, B.E., Hart, R.G., Davies, D.R., Phillips, D.C., and Shore, V.C. 1960. Structure of myoglobin: A three-dimensional Fourier synthesis at 2 Å resolution. *Nature* **185**: 422–427.
- Matthews, B.W. 1966. The determination of the position of the anomalously scattering heavy atom groups in protein crystals. *Acta Crystallogr.* **20**: 230–239.
- Pornillos, O., Chen, Y.-J., Chen, A.P., and Chang, G. 2005. X-ray structure of the EmrE multidrug transporter in complex with a substrate. *Science* **310**: 1950–1953.